

# Estimation of a Component of Variance Due to Competition

BU-134-M

Walter T. Federer\*

May, 1961

## ABSTRACT

A linear model involving competition effects among the individuals within an experimental unit (or stratum) has been defined. The competition effects are defined so that their sum within an experimental unit equals zero, ie., if one or more individuals yields high because they are better competitors the remaining individuals yield low by exactly the same amount. In addition to the competition effect, a random error effect and an experimental unit effect are postulated. By varying the stratum or experimental unit size and by postulating a relationship between a component of variance and size of experimental unit it is possible to estimate a component of variance due to competition as well as the components of variance due to the random error component and the variation among experimental units.

---

\*Biometrics Unit, Plant Breeding Department, Cornell University

## ESTIMATION OF A COMPONENT OF VARIANCE DUE TO COMPETITION

BU-134-M

Walter T. Federer \*

May, 1961

Some theory on a model and a component of variance due to competition is given by Yates and Zacopanay (J. Agric. Sci. 25:545-577, 1935). Little additional theory appears to have been published since their results in this classical paper on plot sampling. The purpose of this note is to present their results in more detail together with an alternative model and a sampling procedure allowing estimation of the component of variance due to competition.

To recapitulate and to amplify the results of Yates and Zacopanay, consider that there are  $m$  experimental units (e.u.) each composed of  $h$  sampling units (s.u.) and that  $k$  of the  $h$  sampling units are randomly selected from each experimental unit. The analysis of variance table for the above is of the form:

Source of variation	d.f.	Mean square	Expected value of mean square
Among e.u.	$m-1$	$A$	$V_s' + kV_p'$
Among s.u. within e.u.	$m(k-1)$	$B$	$V_s'$
Total (corrected for mean)	$mk-1$	-	-

Suppose that the linear equation for yield,  $Y_{ij}$ , is of the form:

$$Y_{ij} = p_i + s_{ij} + c_{ij} \quad (1)$$

where the  $p_i$  are e.u. effects which are identically and independently distributed with mean  $\mu$  and variance  $V_p$ , the  $s_{ij}$  are s.u. effects which are identically and independently distributed with mean zero and variance  $V_s$ , and the  $c_{ij}$  are competition effects between s.u. within an e.u. indicating the effect of one or more s.u. on a given s.u. and are identically and independently distributed with mean  $\bar{c}_i$  and variance  $V_c$ . Yates and Zacopanay give the above (without writing down equation (1)) and further state that the mean of  $k$  s.u. out of the  $h$  s.u. is

$$\frac{1}{k} \sum_{j=1}^k c_j = \frac{h-k}{hk} \sum_{j=1}^k c_i - \frac{1}{h} \sum_{j=k+1}^h c_j$$

\*Biometrics Unit, Plant Breeding Department, Cornell University

with a variance equal to

$$V\left(\frac{1}{k} \sum_{j=1}^k c_j\right) = \left[k\left(\frac{h-k}{hk}\right)^2 + \frac{(h-k)}{h^2}\right] V_c = \frac{h-k}{hk} V_c \quad (2)$$

They further state that the within experimental unit variance is equal to

$$E[A] = V_s + V_c = V_s^* \quad (3)$$

and the variance among plot means is

$$\begin{aligned} E\left[\frac{B}{k}\right] &= V_p + \frac{1}{k} V_s + \frac{h-k}{hk} V_c \\ &= \frac{1}{k} (V_s + V_c) + (V_p - \frac{1}{h} V_c) \\ &= \frac{1}{k} V_s^* + V_p^* \end{aligned} \quad (4)$$

At this point we shall alter the Yates and Zacopanay model given by equation (1) and the definition of the competition effects. They state that the  $c_{ij}$  are independently distributed. Since  $h$  is finite, the  $c_{ij}$  are not independent; therefore, the model will be redefined in that the  $c_{ij}$  are said to be identically distributed with mean zero and variance  $\frac{h-1}{h} V_c = V_c^*$ . This would correspond now to equation (2) with  $k = \text{unity}$ . Otherwise, our model is the same as the one given by Yates and Zacopanay. For  $h$  large, the two models are the same, but this is not true for  $h$  small (e.g.,  $h = \text{litter size} = \text{a small number}$ ).

For sake of generality, let the  $k_i$  vary from experimental unit to experimental unit, but let  $h$ , the size of the e.u., be constant. Then the analysis of variance is of the form:

Source of variation	d.f.	Mean square	Expected value of mean square
Among e.u.	$m-1$	A	$V_s + aV_c + bV_p$
Among s.u. within e.u.	$k_s - m$	B	$V_s + V_c$
Total (corrected for mean)	$k_s - 1$	-	-

where

$$a = (m-1 - \frac{k_{\cdot}}{h} + \frac{\sum k_i^2}{hk_{\cdot}}) / (m-1) ;$$

$$k_{\cdot} = \sum_{i=1}^m k_i ;$$

$$b = (k_{\cdot} - \sum k_i^2 / k_{\cdot}) / (m-1) .$$

When  $k_i$  is a constant  $a=(h-k)/h$  and  $b=k$ . Also, when  $k_i=h$ ,  $a = \text{zero}$  and  $b=h$ .

Since there are three unknown parameters in the two mean squares A and B, no solution is possible unless one of the parameters is known or there is a known relationship between at least two of the three parameters,  $V_c$ ,  $V_s$  and  $V_p$ . Suppose, however, that  $n$  analyses of variance are available for varying values of  $h$  (e.g., say  $h=4,5,\dots,n+4$ ) and that there is a known relationship among the  $V_{ph}$ , among the  $V_{sh}$ , and among the  $V_{ch}$ . For example, it might be reasonable to postulate a polynomial relationship among the variance components for varying  $h$ . If the relationship were quadratic for one of the components, say  $V_{ch}$ , then

$$V_{ch} = \alpha + \beta h + \gamma h^2 + \text{error} \quad (5)$$

and if the relationship were cubic, then

$$V_{ch} = \alpha + \beta h + \gamma h^2 + \delta h^3 + \text{error} \quad (6)$$

Similar relationships (or others) may be postulated from biological, physical, or sociological theory for the  $V_{ph}$  and the  $V_{sh}$ . With these postulated relationships and with  $n$  values of  $h$ , least squares estimates of the regression coefficients ( $\alpha$ ,  $\beta$ ,  $\gamma$ , etc.) in the postulated equations, as in (5) or (6), may be obtained by minimizing the sums of squares of differences between the  $2n$  mean squares and their expected values. Then, the estimated components of variance for each  $h$  may be obtained from standard regression equations, thus:

$$\hat{V}_{ch} = \hat{\alpha} + \hat{\beta}h + \hat{\gamma}h^2 + \hat{\delta}h^3 + \dots \quad (7)$$

$$\hat{V}_{sh} = \hat{a} + \hat{b}h + \hat{c}h^2 + \hat{d}h^3 + \dots \quad (8)$$

$$\hat{V}_{ph} = \hat{A} + \hat{B}h + \hat{C}h^2 + \hat{D}h^3 + \dots \quad (9)$$

where the letter with the hat (^) over it denotes the least squares estimate of the regression coefficient and the ... notation mean that additional polynomial terms are required if the polynomial regression is higher ordered than three.

With  $\delta$ ,  $d$ ,  $D$ , and all higher regression coefficients equal to zero in equations (7), (8), and (9), there are nine parameters to estimate (i.e.,  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $a$ ,  $b$ ,  $c$ ,  $A$ ,  $B$ , and  $C$ ). Consequently,  $h$  must take on at least 5 different values in order to estimate the nine parameters. If  $\gamma$ ,  $c$ , and  $C$  are also zero, then a minimum of 3 values of  $h$  is required in order to estimate the six parameters ( $\alpha$ ,  $\beta$ ,  $a$ ,  $b$ ,  $A$ , and  $B$ ).

Since the variance components are estimated from an ordinary regression equation, the variance formula for a predicted value for ordinary regression may be used to estimate the variance of the estimated variance components. Since the assumption of normality would not be tenable for variance components, tests of significance and confidence intervals cannot be made. If  $m$  is large, the normal approximation may not be too bad.

A component of variance due to competition is useful in many fields. For example, in animal husbandry, estimates of the competition component of variance could be estimated from animals born in a litter or fed together in a feeding trial with specified litter sizes or number of feeding plots. Then, the average weight could be compared with the estimated component of variance due to competition to determine optimum litter size to obtain maximum weight.

In plant studies, the plot size,  $h$ , could be varied and the competition component of variance among plants (equally spaced) within the experimental unit could be estimated. An extension of the model would be required for corn plants planted in hills. Here there would be competition between hills and competition between plants within a hill.

Full use of this component of variance due to competition requires much study. However, enough study has been done for animal husbandry experiments to indicate that estimates of the competition component of variance are potentially useful in studying psychological behavior of animals, particularly swine.

The above results, assuming a quadratic relationship, will be applied to birth weights and weaning weights of pigs from several breeds. The data are those collected by the Animal Husbandry Department of Cornell University.